

# REPRESENTING REAL NUMBERS IN A GENERALIZED NUMERATION SYSTEM

EMILIE CHARLIER, MARION LE GONIDEC, AND MICHEL RIGO

**ABSTRACT.** We show how to represent an interval of real numbers in an abstract numeration system built on a language that is not necessarily regular. As an application, we consider representations of real numbers using the Dyck language. We also show that our framework can be applied to the rational base numeration systems.

## 1. INTRODUCTION

In [LR02], P. Lecomte and the third author showed how to represent an interval of real numbers in an abstract numeration system built on a regular language satisfying some suitable conditions. In this paper, we provide a wider framework and we show that their results can be extended to abstract numeration systems built on a language that is not necessarily regular. Our aim is to provide a unified approach for the representation of real numbers in various numeration systems encountered in the literature [AFS08, DT89, LR01, Lot02].

This paper is organized as follows. In the second section, we recall some useful definitions and results from automata theory. In Section 3, we restate the general framework of [LR02]. Then in Section 4, we show that the infinite words obtained as limits of words of a language are exactly the infinite words having all their prefixes in the corresponding prefix closure. In view of this result, we shall consider only abstract numeration systems built on a prefix-closed language to represent the reals. One can notice that usual numeration systems like integer base systems,  $\beta$ -numeration or substitutive numeration systems are all built on prefix-closed languages [DT89, Lot02]. In Section 5, we show how to represent an interval  $[s_0, 1]$  of real numbers in a generalized abstract numeration system built on a language satisfying some general hypotheses. Finally, in Section 6, we give three applications of our methods, that were not settled yet by the results of [LR02]. First, we consider a non-regular language  $L$  such that its prefix-language  $\text{Pref}(L)$  is regular. In a second part, we consider the representation of real numbers in the generalized abstract numeration system built on the language of the prefixes of the Dyck words. In this case, neither the Dyck language  $D$  nor its prefix-closure  $\text{Pref}(D)$  are recognized by a finite automaton. We compute the complexity functions of this language, i.e., for each word  $w$ , the function mapping an integer  $n$  onto  $\text{Card}(w^{-1}D \cap \{a, b\}^n)$ , and we show that we can apply our results to the corresponding abstract numeration system. The third application that we consider is the abstract numeration system built on the language  $L_{\frac{3}{2}}$  recently introduced in [AFS08]. We show that our method leads, up to some scaling factor, to the same representation of the reals as the one given in [AFS08].

## 2. PRELIMINARIES

Let us recall some usual definitions. For more details, see for instance [Eil74] or [Sak03]. An *alphabet* is a non-empty finite set of symbols, called *letters*. A *word* over an alphabet  $\Sigma$  is a finite or infinite sequence of letters in  $\Sigma$ . The *empty word*

is denoted by  $\varepsilon$ . The set of finite (resp. infinite) words over  $\Sigma$  is denoted by  $\Sigma^*$  (resp.  $\Sigma^\omega$ ). The set  $\Sigma^*$  is the free monoid generated by  $\Sigma$  with respect to the concatenation product of words and with  $\varepsilon$  as neutral element. A *language* (resp.  $\omega$ -language) over  $\Sigma$  is a subset of  $\Sigma^*$  (resp.  $\Sigma^\omega$ ). If  $w$  is a finite word over  $\Sigma$ , the length of  $w$ , denoted by  $|w|$ , is the number of its letters and if  $a \in \Sigma$ , then  $|w|_a$  is the number of occurrences of  $a$  in  $w$ . If  $w$  is a finite (resp. infinite) word over  $\Sigma$ , then for all  $i \in \llbracket 0, |w| - 1 \rrbracket$  (resp.  $i \in \mathbb{N}$ ),  $w[i]$  denotes its  $(i + 1)$ st letter, for all  $0 \leq i \leq j \leq |w| - 1$  (resp.  $0 \leq i \leq j$ ), the *factor*  $w[i, j]$  of  $w$  is the word  $w[i] \cdots w[j]$ , and for all  $i \in \llbracket 0, |w| \rrbracket$  (resp.  $i \in \mathbb{N}$ ),  $w[0, i - 1]$  is the *prefix* of length  $i$  of  $w$ , where we set  $w[0, -1] := \varepsilon$ . The set of prefixes of a word  $w$  (resp. a language  $L$ ) is denoted by  $\text{Pref}(w)$  (resp.  $\text{Pref}(L)$ ). Notice that indices are counted from 0.

One can endow  $\Sigma^\omega \cup \Sigma^*$  with a metric space structure as follows. If  $x$  and  $y$  are two distinct infinite words over  $\Sigma$ , define the distance  $d$  over  $\Sigma^\omega$  by  $d(x, y) := 2^{-\ell}$  where  $\ell = \inf\{i \in \mathbb{N} \mid x[i] \neq y[i]\}$  is the length of the maximal common prefix between  $x$  and  $y$ . We set  $d(x, x) = 0$  for all  $x \in \Sigma^\omega$ . This distance can be extended to  $\Sigma^\omega \cup \Sigma^*$  by replacing the finite words  $z$  by  $z\#^\omega$ , where  $\#$  is a new letter not in  $\Sigma$ . A sequence  $(w^{(n)})_{n \geq 0}$  of words over  $\Sigma$  *converges* to an infinite word  $w$  over  $\Sigma$  if  $d(w^{(n)}, w) \rightarrow 0$  as  $n \rightarrow +\infty$ .

A deterministic (finite or infinite) automaton over an alphabet  $\Sigma$  is a directed graph  $\mathcal{A} = (Q, q_0, \Sigma, \delta, F)$ , where  $Q$  is the set of *states*,  $q_0$  is the *initial state*,  $F \subseteq Q$  is the set of *final states* and  $\delta: Q \times \Sigma \rightarrow Q$  is the *transition function*. The transition function can be naturally extended to  $Q \times \Sigma^*$  by  $\delta(q, \varepsilon) = q$  and  $\delta(q, aw) = \delta(\delta(q, a), w)$  for all  $q \in Q$ ,  $a \in \Sigma$  and  $w \in \Sigma^*$ . We often use  $q \cdot w$  as shorthand for  $\delta(q, w)$ . A state  $q \in Q$  is *accessible* (resp. *coaccessible*) if there exists a word  $w \in \Sigma^*$  such that  $\delta(q_0, w) = q$  (resp.  $\delta(q, w) \in F$ ) and  $\mathcal{A}$  is *accessible* (resp. *coaccessible*) if all its state are accessible (resp. coaccessible). A word  $w \in \Sigma^*$  is *accepted* by  $\mathcal{A}$  if  $\delta(q_0, w) \in F$ . The set of accepted words is the *language recognized by  $\mathcal{A}$* . A deterministic automaton is said to be *finite* (resp. *infinite*) if its set of states is finite (resp. infinite). A language is *regular* if it is recognized by some deterministic finite automaton (DFA).

Among all the deterministic automata recognizing a language, one can distinguish the minimal automaton of this language, which is unique up to isomorphism and is defined as follows. The *minimal automaton* of a language  $L$  over an alphabet  $\Sigma$  is the deterministic automaton  $\mathcal{A}_L = (Q_L, q_{0,L}, \Sigma, \delta_L, F_L)$  where the states are the sets  $w^{-1}L = \{x \in \Sigma^* \mid wx \in L\}$ , for any  $w \in \Sigma^*$ , the initial state is  $q_{0,L} = \varepsilon^{-1}L = L$ , the final states are the sets  $w^{-1}L$  with  $w \in L$  and the transition function  $\delta_L$  is defined by  $\delta_L(w^{-1}L, a) = (wa)^{-1}L$  for all  $w \in \Sigma^*$  and all  $a \in \Sigma$ . By construction,  $\mathcal{A}_L$  is accessible and the set of accepted words is exactly  $L$ . It is well known that  $\mathcal{A}_L$  is finite if and only if  $L$  is regular. The *trim minimal automaton* of a language is the minimal automaton of this language from which the only possible sink state has been removed, i.e. we keep only the coaccessible states. In this case, the transition function can possibly be a partial function.

If  $L$  is the language recognized by a deterministic automaton  $\mathcal{A} = (Q, q_0, \Sigma, \delta, F)$ ,  $L_q := \{w \in \Sigma^* \mid \delta(q, w) \in F\}$  is the language of the words accepted from the state  $q$  in  $\mathcal{A}$  and  $u_q(n)$  (resp.  $v_q(n)$ ) is the number of words of length  $n$  (resp. less or equal to  $n$ ) in  $L_q$ . The maps  $u_q: \mathbb{N} \rightarrow \mathbb{N}$  are called the *complexity functions of  $\mathcal{A}$* . The language  $L$  is *polynomial* if  $u_{q_0}(n)$  is  $\mathcal{O}(n^k)$  for some non-negative integer  $k$  and *exponential* if  $u_{q_0}(n)$  is  $\Omega(\theta^n)$  for some  $\theta > 1$ , i.e., if there exists a constant  $c > 0$  such that  $u_{q_0}(n) \geq c\theta^n$  for infinitely many non-negative integers  $n$ .

### 3. GENERALIZED ABSTRACT NUMERATION SYSTEMS

If  $L$  is a language over a totally ordered alphabet  $(\Sigma, <)$ , the *genealogical (or radix) ordering*  $<_{gen}$  over  $L$  induced by  $<$  is defined as follows. The words of the language are ordered by increasing length and for words of the same length, one uses the lexicographical ordering induced by  $<$ . Recall that for two words  $x, y \in \Sigma^*$  of same length,  $x$  is lexicographically less than  $y$  if there exist  $w, x', y' \in \Sigma^*$  and  $a, b \in \Sigma$  such that  $x = wax'$ ,  $y = wby'$  and  $a < b$ . The lexicographical ordering is naturally extended to infinite words.

**Definition 1.** A (generalized) abstract numeration system is a triple  $S = (L, \Sigma, <)$  where  $L$  is an infinite language over a totally ordered alphabet  $(\Sigma, <)$ . Enumerating the words of  $L$  using the genealogical order  $<_{gen}$  induced by the ordering  $<$  on  $\Sigma$  gives a one-to-one correspondence  $\text{rep}_S: \mathbb{N} \rightarrow L$  mapping the non-negative integer  $n$  onto the  $(n+1)$ st word in  $L$ . In particular, 0 is sent onto the first word in the genealogically ordered language  $L$ . The reciprocal map is denoted by  $\text{val}_S: L \rightarrow \mathbb{N}$  and for all  $w \in L$ ,  $\text{val}_S(w)$  is called the  $S$ -numerical value of  $w$ .

Compare with [LR01], we do not ask the language of the numeration to be regular. It is the reason for the introduction of the terminology “generalized”.

**Example 2.** Let  $\Sigma = \{a, b\}$ ,  $L = \{w \in \Sigma^*: ||w|_a - |w|_b| \leq 1\}$ , and  $S = (L, \Sigma, a < b)$ . The minimal automaton of  $L$  is given in Figure 1. The first words of the  $L$  are

$\varepsilon, a, b, ab, ba, aab, aba, abb, baa, bab, bba, aabb, abab, abba, baab, \dots$

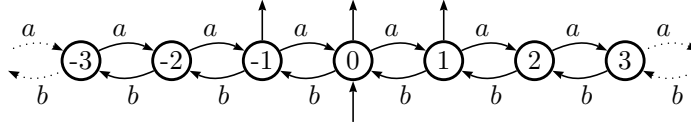


FIGURE 1. The minimal automaton of  $L$ .

The following proposition is a result from [LR02] extended to any language. This shows how to compute the numerical value of a word in the numeration language.

**Proposition 3.** Let  $S = (L, \Sigma, <)$  be a (generalized) abstract numeration system and let  $\mathcal{A} = (Q, q_0, \Sigma, \delta, F)$  be a deterministic automaton recognizing  $L$ . If  $w \in L$ , then we have

$$\text{val}_S(w) = v_{q_0}(|w| - 1) + \sum_{i=0}^{|w|-1} \sum_{a < w[i]} u_{q_0 \cdot w[0, i-1]a}(|w| - i - 1).$$

### 4. LANGUAGES $L$ WITH UNCOUNTABLE $\text{Adh}(L)$

The notion of adherence has been introduced in [Niv78] and has been extensively studied in [BN80].

**Definition 4.** Let  $L$  be a language over an alphabet  $\Sigma$ . The *adherence* of  $L$ , denoted by  $\text{Adh}(L)$ , is the set of infinite words over  $\Sigma$  whose prefixes are prefixes of words in  $L$ :

$$\text{Adh}(L) = \{w \in \Sigma^\omega \mid \text{Pref}(w) \subseteq \text{Pref}(L)\}.$$

Notice that  $\text{Adh}(L)$  is empty if and only if  $L$  is finite.

For the usual topology on  $\Sigma^* \cup \Sigma^\omega$ , the closure  $\bar{L}$  of a language  $L$  over  $\Sigma$  satisfies the equality:  $\bar{L} = L \cup \text{Adh}(L)$ .

The following lemma gives a characterization of the adherence of a language [BN80]. We give a proof for the sake of completeness.

**Lemma 5.** *Let  $L$  be a language over an alphabet  $\Sigma$ . The adherence of  $L$  is the set of infinite words over  $\Sigma$  that are limits of words in  $L$ :*

$$\text{Adh}(L) = \{w \in \Sigma^\omega \mid \exists (w^{(n)})_{n \geq 0} \in L^\mathbb{N}, w^{(n)} \rightarrow w\}.$$

*Proof.* Take an infinite word  $w$  in  $\text{Adh}(L)$ . Then for all  $n \geq 0$ , we have  $w[0, n-1] \in \text{Pref}(L)$ . Thus for all  $n \geq 0$ , there exists a finite word  $z^{(n)} \in \Sigma^*$  such that  $w^{(n)} := w[0, n-1]z^{(n)}$  belongs to  $L$ . Obviously  $w^{(n)} \rightarrow w$  and  $w$  belongs to the r.h.s. set in the statement. Conversely, take an infinite word  $w$  which is the limit of a sequence  $(w^{(n)})_{n \geq 0}$  of words in  $L$ . Then for all  $\ell \geq 0$ , there exists  $n \geq 0$  such that we have  $w[0, \ell-1] \in \text{Pref}(w^{(n)}) \subseteq \text{Pref}(L)$ . This shows that  $w$  belongs to  $\text{Adh}(L)$ .  $\square$

The notion of center of a language can be found in [BN80].

**Definition 6.** Let  $L$  be a language over an alphabet  $\Sigma$ . The *center* of  $L$ , denoted by  $\text{Center}(L)$ , is the prefix-closure of the adherence of  $L$ :

$$\text{Center}(L) = \text{Pref}(\text{Adh}(L)).$$

The next lemma gives a characterization of the center of a language [BN80]. Again we give a proof for the sake of completeness.

**Lemma 7.** *Let  $L$  be a language over an alphabet  $\Sigma$ . The center of  $L$  is the set of words which are prefixes of an infinite number of words in  $L$ :*

$$\text{Center}(L) = \{w \in \text{Pref}(L) \mid w^{-1}L \text{ is infinite}\}.$$

*Proof.* Take a word  $w$  in  $\text{Center}(L)$ . By definition, there exists a infinite word  $z$  over  $\Sigma$  such that  $wz$  belongs to  $\text{Adh}(L)$ . Then for all  $n \geq 0$ ,  $wz[0, n-1]$  belongs to  $\text{Pref}(L)$ . Thus for all  $n \geq 0$ , there exists a finite word  $y^{(n)} \in \Sigma^*$  such that  $w^{(n)} := wz[0, n-1]y^{(n)}$  belongs to  $L$ , and there are infinitely many such words  $w^{(n)}$ . Conversely, let  $w$  be a prefix of infinitely many words in  $L$ . There exists a letter  $a \in \Sigma$  such that  $wa$  is a prefix of infinitely many words in  $L$ . Iterating this argument, there exists a sequence  $(a_n)_{n \geq 0}$  of letters in  $\Sigma$  such that  $wa_0 \cdots a_n$  belongs to  $\text{Pref}(L)$  for all  $n \geq 0$ . This implies that  $wa_0a_1 \cdots$  belongs to  $\text{Adh}(L)$ . Hence  $w$  belongs to  $\text{Center}(L)$ .  $\square$

**Definition 8.** If  $L$  is a language over an alphabet  $\Sigma$ ,

$$L_\infty = \{w \in \Sigma^\omega \mid \exists^\infty n \in \mathbb{N}, w[0, n-1] \in L\}$$

denotes the set of infinite words over  $\Sigma$  having infinitely many prefixes in  $L$ .

Again, observe that  $L_\infty$  is empty if and only if  $L$  is finite.

The following lemma is obvious.

**Lemma 9.** *For any language  $L$ , we have  $L_\infty \subseteq \text{Adh}(L)$ . Moreover, if  $L$  is a prefix-closed language, then  $L_\infty = \text{Adh}(L)$ .*

Let us recall two results from [LR02].

**Proposition 10.** *Let  $L$  be a regular language. The set  $\text{Adh}(L)$  is uncountably infinite if and only if, in any deterministic finite automaton accepting  $L$ , there exist at least two distinct cycles  $(p_1, \dots, p_r, p_1)$  and  $(q_1, \dots, q_s, q_1)$  where  $r, s \geq 2$ , starting from the same accessible and coaccessible state  $p_1 = q_1$ .*

**Proposition 11.** *Let  $L$  be a regular language. The set  $L_\infty$  is uncountably infinite if and only if, in any deterministic finite automaton accepting  $L$ , there exist at least two distinct cycles  $(p_1, \dots, p_r, p_1)$  and  $(q_1, \dots, q_s, q_1)$  where  $r, s \geq 2$ , starting from the same accessible state  $p_1 = q_1$  and such that each of them contains at least a final state.*

It is well known [SYZS92] that the set of regular languages splits into two parts: the set of exponential languages and the set of polynomial languages. The polynomial regular languages over an alphabet  $\Sigma$  are exactly those that are finite union of languages of the form

$$(1) \quad x_1 y_1^* x_2 y_2^* \cdots x_k y_k^* x_{k+1}$$

where  $k \geq 0$  and the  $x_i$ 's and the  $y_i$ 's are finite words over  $\Sigma$ . Consequently, in view of Proposition 10, the following result is obvious.

**Corollary 12.** *If  $L$  is a regular language, then the following assertions are equivalent:*

- $\text{Adh}(L)$  is an uncountable set;
- $L$  is exponential;
- $\text{Pref}(L)$  is exponential.

If the considered language is not regular, then only the sufficient conditions of Proposition 10 and Proposition 11 hold true. They can be reexpressed as follows.

**Proposition 13.** *If, in any deterministic automaton accepting a language  $L$ , there exist at least two distinct cycles  $(p_1, \dots, p_r, p_1)$  and  $(q_1, \dots, q_s, q_1)$  where  $r, s \geq 2$ , starting from the same accessible and coaccessible state  $p_1 = q_1$ , then the set  $\text{Adh}(L)$  is uncountably infinite and  $L$  is exponential.*

**Proposition 14.** *If, in any deterministic automaton accepting a language  $L$ , there exist at least two distinct cycles  $(p_1, \dots, p_r, p_1)$  and  $(q_1, \dots, q_s, q_1)$  where  $r, s \geq 2$ , starting from the same accessible state  $p_1 = q_1$  and such that each of them contains at least a final state, then the set  $L_\infty$  is uncountably infinite and  $L$  is exponential.*

There exist non-regular exponential languages with an uncountable associated set  $L_\infty$ , and thus also with an uncountable set  $\text{Adh}(L)$ , that are recognized by a deterministic automaton without distinct cycles satisfying condition of Proposition 13. For instance, see Example 43 of Section 6 about the  $\frac{3}{2}$ -number system. Notice that the corresponding trim minimal automaton depicted in Figure 6 has an infinite number of final states. Note that, by considering automata having a finite set of final states, we get back the necessary condition of Proposition 11.

**Proposition 15.** *Let  $L$  be a language recognized by a deterministic automaton  $\mathcal{A}$  having a finite set of final states. The set  $L_\infty$  is uncountably infinite if and only if there exist in  $\mathcal{A}$  at least two distinct cycles  $(p_1, \dots, p_r, p_1)$  and  $(q_1, \dots, q_s, q_1)$  where  $r, s \geq 2$ , starting from the same accessible state  $p_1 = q_1$  and such that each of them contains at least a final state.*

*Proof.* In view of Proposition 13, we only have to show that the condition is necessary. Since there is only a finite number of final states, if  $w \in L_\infty$ , then there exist a final state  $f$  and infinitely many  $n$  such that  $q_0 \cdot w[0, n-1] = f$ . If  $\mathcal{A}$  does not contain such distinct cycles, then this implies that any word in  $L_\infty$  is of the form  $xy^\omega$ , where  $x, y$  are finite words. Since there is a countable number of such words, we would get that  $L_\infty$  is a countable set. The conclusion follows.  $\square$

**Corollary 16.** *Let  $L$  be a language recognized by a deterministic automaton  $\mathcal{A}$  having a finite set of final states. If  $L_\infty$  is an uncountable set, then  $L$  is exponential.*

**Remark 17.** Any deterministic automaton recognizing a non-regular prefix-closed language has an infinite number of final states. Indeed, in such an automaton, all coaccessible states are final.

There exist exponential (and prefix-closed) languages  $L$  with a countable, and even finite, set  $\text{Adh}(L)$ . We give an example of such a language.

**Example 18.** Let  $L = \{w \in \{a, b\}^* \mid \exists u \in \{a, b\}^* : w = a^{\lfloor \frac{|w|}{2} \rfloor} u\}$ . We have

$$u_L(n) = \begin{cases} 2^{\frac{n}{2}} & \text{if } n \equiv 0 \pmod{2}, \\ 2^{\frac{n+1}{2}} & \text{if } n \equiv 1 \pmod{2} \end{cases}$$

and  $\text{Adh}(L) = L_\infty = \{a^\omega\}$ . The minimal automaton of  $L$  is depicted in Figure 2.

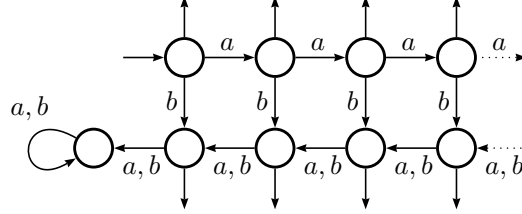


FIGURE 2. The minimal automaton of  $L$ .

## 5. REPRESENTATION OF REAL NUMBERS

In the framework of [LR02], a real number is represented in an abstract numeration system built on a regular language  $L$  as a limit of a sequence of words of  $L$ . Observe that in this context, thanks to Lemma 5, the set of possible representations of the considered reals is  $\text{Adh}(L)$ . Therefore, one could consider abstract numeration systems built on the prefix-language instead of the one built on the language itself, see Remark 20 and Remark 21. This point of view is relevant if we compare this with the framework of the classical integer base  $b \geq 2$  numeration systems. Indeed, in these systems, the numeration language is

$$\mathcal{L}_b := \{1, 2, \dots, b-1\}\{0, 1, \dots, b-1\}^*,$$

which is of course a prefix-closed language. Notice that this is also the case for non-standard numeration systems like  $\beta$ -numeration systems and substitutive numeration systems. Adopting this new framework, we consider only abstract numeration systems built on prefix-closed languages. Therefore, to represent real numbers, we do not distinguish anymore abstract numeration systems built on two distinct languages  $L$  and  $M$  such that  $\text{Pref}(L) = \text{Pref}(M)$ .

Let  $S = (L, \Sigma, <)$  be a generalized abstract numeration system built on a prefix-closed language  $L$ . Let  $\mathcal{A} = (Q, q_0, \Sigma, \delta, F)$  be an accessible deterministic automaton recognizing  $L$ . We make the following assumptions:

### Hypotheses.

- (H1) The set  $\text{Adh}(L)$  is uncountable;
- (H2)  $\forall w \in \Sigma^*, \exists r_w \geq 0: \lim_{n \rightarrow +\infty} \frac{u_{q_0 \cdot w}(n - |w|)}{v_{q_0}(n)} = r_w$ ;
- (H3)  $\forall w \in \text{Adh}(L), \lim_{\ell \rightarrow +\infty} r_{w[0, \ell-1]} = 0$ .

Observe that for all  $w \notin \text{Center}(L)$ , we have  $r_w = 0$ .

Recall that, since  $L$  is a prefix-closed language, we have  $\text{Adh}(L) = L_\infty$ , see Lemma 9.

**Notation.** We set  $r_0 := r_\varepsilon$  and

$$s_0 := 1 - r_0 = \lim_{n \rightarrow +\infty} \frac{v_{q_0}(n-1)}{v_{q_0}(n)}.$$

**Remark 19.** In [LR02] are considered regular languages  $L$  with uncountably infinite  $\text{Adh}(L)$  such that, for each state  $q$  of a DFA recognizing  $L$ , either  $L_q$  is finite, or  $u_q(n) \sim P_q(n)\theta_q^n$  where  $P_q \in \mathbb{R}[X]$  and  $\theta_q \geq 1$ . One can notice that such languages

satisfy the hypotheses (H1), (H2) and (H3) above. Indeed, for all states  $q$  and all  $\ell \geq 0$ , it can be shown that

$$\lim_{n \rightarrow +\infty} \frac{u_q(n - \ell)}{v_{q_0}(n)} = \frac{a_q(\theta_{q_0} - 1)}{\theta_{q_0}^{\ell+1}}$$

where  $\theta_{q_0} > 1$  and  $a_q := \lim_{n \rightarrow +\infty} \frac{u_q(n)}{u_{q_0}(n)}$ . Since  $Q$  is finite, this is sufficient to verify our assumptions. Notice also that for the integer base  $b$  numeration system, the three hypotheses are trivially satisfied.

We shall represent real numbers by infinite words  $w$  of  $\text{Adh}(L)$  by considering the corresponding limit

$$(2) \quad \lim_{n \rightarrow +\infty} \frac{\text{val}_S(w[0, n-1])}{v_{q_0}(n)}.$$

Our aim is to show that for all  $w \in \text{Adh}(L)$ , the limit (2) exists, see Proposition 26.

**Remark 20.** If the considered abstract numeration system is built on a language that is not prefix-closed, we cannot guarantee that the limit (2) exists. Consider for instance the abstract numeration system built on the language  $L$  of Example 2, which is not prefix-closed. The sequences  $((ab)^n)_{n \geq 0}$  and  $((ab)^n a)_{n \geq 0}$  of words in  $L$  converge to the same infinite word  $(ab)^\omega$ , but the corresponding numerical sequences do not converge to the same real number. More precisely, using notation of Example 2, we have

$$(3) \quad \lim_{n \rightarrow +\infty} \frac{\text{val}_S((ab)^n)}{v_0(2n)} = \frac{3}{4} \quad \text{and} \quad \lim_{n \rightarrow +\infty} \frac{\text{val}_S((ab)^n a)}{v_0(2n+1)} = \frac{3}{5},$$

so that the limit

$$\lim_{n \rightarrow +\infty} \frac{\text{val}_S((ab)^\omega[0, n-1])}{v_0(n)}$$

does not exist. This essentially comes from the staircase behaviour of  $(u_0(n))_{n \geq 0}$ . We have that for all  $n \geq 0$ ,

$$u_0(n) = \begin{cases} \binom{n}{\frac{n}{2}} & \text{if } n \equiv 0 \pmod{2}, \\ 2\binom{n-1}{\frac{n-1}{2}} & \text{if } n \equiv 1 \pmod{2}. \end{cases}$$

This implies in particular that  $\lim_{n \rightarrow +\infty} \frac{v_0(n-1)}{v_0(n)}$  does not exist. Indeed, using Stirling formula and [Bou07, Ch. V.4, Prop. 2], we have

$$(4) \quad v_0(2n) \sim \frac{8}{3\sqrt{\pi}} n^{-\frac{1}{2}} 4^n \quad \text{and} \quad v_0(2n-1) \sim \frac{5}{3\sqrt{\pi}} n^{-\frac{1}{2}} 4^n \quad (n \rightarrow +\infty).$$

Hence,

$$\lim_{n \rightarrow +\infty} \frac{v_0(2n-1)}{v_0(2n)} = \frac{5}{8} \quad \text{and} \quad \lim_{n \rightarrow +\infty} \frac{v_0(2n)}{v_0(2n+1)} = \frac{2}{5}.$$

By Proposition 3, we obtain that for all  $n \geq 1$ ,

$$\begin{aligned} \frac{\text{val}_S((ab)^n)}{v_0(2n)} &= \frac{v_0(2n-1)}{v_0(2n)} + \frac{\sum_{i=0}^{n-1} u_2(2i)}{v_0(2n)}, \\ \frac{\text{val}_S((ab)^n a)}{v_0(2n+1)} &= \frac{v_0(2n)}{v_0(2n+1)} + \frac{\sum_{i=0}^{n-1} u_2(2i+1)}{v_0(2n+1)}. \end{aligned}$$

Using again Stirling formula, we get

$$u_2(2i) = \binom{2i}{i-1} \sim \frac{1}{\sqrt{\pi}} i^{-\frac{1}{2}} 4^i \quad (i \rightarrow +\infty),$$

$$u_2(2i+1) = \binom{2i+1}{i} + \binom{2i+1}{i-1} \sim \frac{4}{\sqrt{\pi}} i^{-\frac{1}{2}} 4^i \quad (i \rightarrow +\infty).$$

Therefore, by [Bou07, Ch. V.4, Prop. 2] and in view (4), it follows that

$$\lim_{n \rightarrow +\infty} \frac{\sum_{i=0}^{n-1} u_2(2i)}{v_0(2n)} = \frac{1}{8} \quad \text{and} \quad \lim_{n \rightarrow +\infty} \frac{\sum_{i=0}^{n-1} u_2(2i+1)}{v_0(2n+1)} = \frac{1}{5}.$$

and we obtain the limits of (3).

**Remark 21.** Considering prefix-closed languages not only avoids numerical convergence problems as in Remark 20 but also permits to get rid of problems arising from languages  $L$  such that there is infinitely many  $n$  for which  $L \cap \Sigma^n = \emptyset$  as discussed in [LR02, Remark 4].

**Definition 22.** If  $w \in \text{Adh}(L)$  is such that  $\lim_{n \rightarrow +\infty} \frac{\text{val}_S(w[0, n-1])}{v_{q_0}(n)} = x$ , we say that  $w$  is an  $S$ -representation of  $x$ .

**Example 23.** Consider the abstract numeration system built on the Dyck language that will be described in Example 42. Table 1 gives some numerical approximations. We will see further that  $\lim_{n \rightarrow +\infty} \frac{\text{val}_S((aab)^\omega[0, n-1])}{v_{q_0}(n)} = \frac{39}{49} = 0.79592 \dots$ .

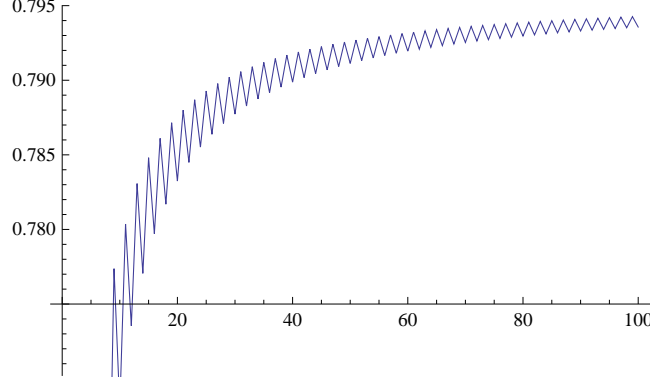
$w$	$\text{val}_S(w)$	$v_{q_0}( w )$	$\frac{\text{val}_S(w)}{v_{q_0}( w )}$
$a$	1	2	0.50000
$aa$	2	4	0.50000
$aab$	5	7	0.71429
$aaba$	9	13	0.69231
$aabaa$	17	23	0.73913
$aabaab$	32	43	0.74419
$aabaaba$	60	78	0.76923
$aabaabaa$	112	148	0.75676
$aabaabaab$	213	274	0.77737
$aabaabaaba$	404	526	0.76806
$aabaabaabaa$	771	988	0.78036
$aabaabaabaab$	1479	1912	0.77354
$aabaabaabaaba$	2841	3628	0.78308
$aabaabaabaabaa$	5486	7060	0.77705
$aabaabaabaabaab$	10591	13495	0.78481
$\vdots$	$\vdots$	$\vdots$	$\vdots$

TABLE 1. Some numerical approximations.

Notice that for all  $w \in \text{Adh}(L)$ , we have  $\text{val}_S(w[0, n-1]) \in [v_{q_0}(n-1), v_{q_0}(n)-1]$  for all  $n \geq 1$ . Therefore, the represented real numbers  $x$  must belong to the interval  $[s_0, 1]$ .

Like in [LR02], we divide  $[s_0, 1]$  into subintervals  $I_y$ , for all prefixes  $y$  of infinitely many words in  $L$ . For each  $\ell \geq 0$ ,  $\text{Center}(L) \cap \Sigma^\ell$  is the set of words of length  $\ell$  which are prefixes of infinitely many words of  $L$ . For each  $y \in \text{Center}(L) \cap \Sigma^\ell$  and




 FIGURE 3. The first 100 values of  $\frac{\text{val}_S((aab)^\omega[0, n-1])}{v_{q_0}(n)}$ .

$n \geq \ell \geq 0$ , define

$$\alpha_{y,n} := \frac{v_{q_0}(n-1)}{v_{q_0}(n)} + \sum_{\substack{x < y \\ x \in \text{Center}(L) \cap \Sigma^\ell}} \frac{u_{q_0 \cdot x}(n-\ell)}{v_{q_0}(n)}$$

and

$$I_{y,n} := \left[ \alpha_{y,n}, \alpha_{y,n} + \frac{u_{q_0 \cdot y}(n-\ell)}{v_{q_0}(n)} \right].$$

Then, in view of Hypothesis (H2), for all  $y \in \text{Center}(L) \cap \Sigma^\ell$ , we can define the limit interval

$$I_y := \lim_{n \rightarrow +\infty} I_{y,n} = [\alpha_y, \alpha_y + r_y],$$

where

$$\alpha_y := \lim_{n \rightarrow +\infty} \alpha_{y,n} = s_0 + \sum_{\substack{x < y \\ x \in \text{Center}(L) \cap \Sigma^\ell}} r_x.$$

Moreover, we set  $I_y := \emptyset$  for all  $y \in L \setminus \text{Center}(L)$ . From [LR02], we know that for all  $\ell \geq 0$ , we have

$$[s_0, 1] = \bigcup_{y \in \text{Center}(L) \cap \Sigma^\ell} I_y$$

and for all  $y, z \in \Sigma^*$ ,

$$(5) \quad I_{yz} \subseteq I_y.$$

More precisely, if  $a_1, \dots, a_k$  are the letters of  $\Sigma$  and if  $a_1 < \dots < a_k$ , then for all  $y \in \text{Center}(L)$  and all  $j \in \llbracket 1, k \rrbracket$  such that  $ya_j \in \text{Center}(L)$ , one has

$$(6) \quad I_{ya_j} = \left[ \alpha_y + \sum_{i=1}^{j-1} r_{ya_i}, \alpha_y + \sum_{i=1}^j r_{ya_i} \right].$$

**Remark 24.** Let  $y, z$  be words in  $\Sigma^*$  such that  $yz \in L$ . If  $y$  is prefix of infinitely many words in  $L$  and if  $|z|$  is large enough so that every word of length  $|yz|$  has a prefix in  $\text{Center}(L) \cap \Sigma^{|y|}$ , then we have

$$(7) \quad \text{val}_S(yz) = v_{q_0}(|yz|-1) + \sum_{\substack{x < y \\ x \in \text{Center}(L) \cap \Sigma^{|y|}}} u_{q_0 \cdot x}(|z|) + \sum_{i=|y|}^{|yz|-1} \sum_{\substack{x < yz[0, i] \\ |x|=i+1}} u_{q_0 \cdot x}(|yz|-i-1).$$

**Lemma 25.** *Let  $w \in \text{Adh}(L)$ . For all  $\ell \geq 0$ ,  $w[0, \ell - 1]$  belongs to  $\text{Center}(L) \cap \Sigma^\ell$  and the limit*

$$\lim_{\ell \rightarrow +\infty} \alpha_{w[0, \ell - 1]}$$

*exists.*

*Proof.* The first part is obvious since  $w[0, \ell - 1]$  is a prefix of  $w[0, n - 1]$  for any  $n \geq \ell$ , see Lemma 7. For the second part, on the one hand, observe that (5) implies that for all  $\ell \geq 1$ ,  $\alpha_{w[0, \ell - 1]} \leq \alpha_{w[0, \ell]}$ . On the other hand, we have also that for all  $\ell \geq 1$ ,  $\alpha_{w[0, \ell - 1]} \leq 1$ . Hence,  $(\alpha_{w[0, \ell - 1]})_{\ell \geq 1}$  is a bounded and non-decreasing sequence, so it must converge.  $\square$

**Notation.** For all  $w \in \text{Adh}(L)$ ,  $\alpha_w := \lim_{\ell \rightarrow +\infty} \alpha_{w[0, \ell - 1]}$ .

Note that we have  $\alpha_w \geq \alpha_{w[0, \ell - 1]}$  for all  $\ell \geq 1$ .

**Proposition 26.** *For all  $w \in \text{Adh}(L)$ , we have*

$$\lim_{n \rightarrow +\infty} \frac{\text{val}_S(w[0, n - 1])}{v_{q_0}(n)} = \alpha_w.$$

*Proof.* Let  $w \in \text{Adh}(L)$ . For all  $\ell$  and  $n$  such that  $n \geq \ell \geq 1$ , we have

$$(8) \quad \alpha_{w[0, \ell - 1], n} \leq \frac{\text{val}_S(w[0, n - 1])}{v_{q_0}(n)} < \alpha_{w[0, \ell - 1], n} + \frac{u_{q_0 \cdot w[0, \ell - 1]}(n - \ell)}{v_{q_0}(n)}.$$

Let  $\varepsilon > 0$ . For all  $\ell \geq 1$ , there exists  $N(\ell) \geq \ell$  such that for all  $n \geq N(\ell)$ , we have

$$\alpha_{w[0, \ell - 1]} - \frac{\varepsilon}{2} < \frac{\text{val}_S(w[0, n - 1])}{v_{q_0}(n)} < \alpha_{w[0, \ell - 1]} + r_{w[0, \ell - 1]} + \frac{\varepsilon}{2}.$$

By Hypothesis (H3) and Lemma 25, there exists also  $k \in \mathbb{N}$  such that for all  $\ell \geq k$ ,

$$r_{w[0, \ell - 1]} < \frac{\varepsilon}{2} \quad \text{and} \quad 0 < \alpha_w - \alpha_{w[0, \ell - 1]} < \frac{\varepsilon}{2}.$$

It follows that for all  $n \geq N(k)$ ,

$$\alpha_w - \varepsilon < \alpha_{w[0, k - 1]} - \frac{\varepsilon}{2} < \frac{\text{val}_S(w[0, n - 1])}{v_{q_0}(n)} < \alpha_w + \varepsilon$$

and the conclusion follows.  $\square$

The preceding proposition allows us to define the  $S$ -value of an infinite word in  $\text{Adh}(L)$ .

**Definition 27.** The application  $\text{val}_S: \text{Adh}(L) \rightarrow [s_0, 1]: w \mapsto \alpha_w$  is called the  $S$ -value function.

**Proposition 28.** *If  $w, z \in \text{Adh}(L)$  are such that  $w$  is lexicographically less than  $z$ , then  $\text{val}_S(w) \leq \text{val}_S(z)$ .*

*Proof.* Let  $w, z \in \text{Adh}(L)$ . We deduce from (6) that if  $k := \inf\{i \in \mathbb{N} \mid w[i] < z[i]\}$ , then  $\forall \ell \geq k$ , we have  $\alpha_{w[0, \ell - 1]} \leq \alpha_{z[0, \ell - 1]}$  and the proposition holds.  $\square$

Recall now a result from [BB97].

**Lemma 29.** *If  $K$  is an infinite language over a totally ordered alphabet, then  $\text{Adh}(K)$  contains a minimal element for the lexicographical ordering.*

This leads to the following definition.

**Definition 30.** For all  $y \in \text{Center}(L)$ ,  $m_y$  (resp.  $M_y$ ) denotes the least (resp. greater) word in  $\text{Adh}(L)$  in the lexicographical ordering having  $y$  as a prefix.

Notice that for all  $y \in \text{Center}(L)$ , we have  $m_y = wv$  (resp.  $M_y = wu$ ), where  $u$  (resp.  $v$ ) is the minimal (resp. maximal) word in  $\text{Adh}(y^{-1}L)$  for the lexicographical ordering.

**Example 31.** Continuing Example 23, we have  $m_{aab} = aaba^\omega$  and  $M_{aab} = aabb(ab)^\omega$ .

**Lemma 32.** *For all  $y \in \text{Center}(L)$ , one has*

$$\text{val}_S(m_y) = \alpha_y \text{ and } \text{val}_S(M_y) = \alpha_y + r_y.$$

*Proof.* Let  $y \in \text{Center}(L)$ . From (6), we get that for all  $\ell \geq |y|$ ,  $\alpha_{m_y[0, \ell-1]} = \alpha_y$  and  $\alpha_{M_y[0, \ell-1]} + r_{M_y[0, \ell-1]} = \alpha_y + r_y$ . Therefore, we obtain that for all  $\ell \geq |y|$ ,

$$\begin{aligned} \alpha_y &\leq \text{val}_S(m_y) \leq \alpha_y + r_{m_y[0, \ell-1]}, \\ \alpha_y + r_y - r_{M_y[0, \ell-1]} &\leq \text{val}_S(M_y) \leq \alpha_y + r_y. \end{aligned}$$

We conclude by using Hypothesis (H3).  $\square$

**Proposition 33.** *The  $S$ -value function is uniformly continuous.*

*Proof.* Let  $w, z \in \text{Adh}(L)$ . Assume that  $d(w, z) = 2^{-\ell}$ . Then  $w[0, \ell-1] = z[0, \ell-1]$  and, in view of Lemma 32, the  $S$ -values  $\text{val}_S(w)$  and  $\text{val}_S(z)$  belong to  $I_{w[0, \ell-1]}$ . Thus  $|\text{val}_S(w) - \text{val}_S(z)| \leq r_{w[0, \ell-1]} \rightarrow 0$  as  $\ell \rightarrow +\infty$  by Hypothesis (H3). The conclusion follows.  $\square$

Using Lemma 32, we are able to give an expression of the  $S$ -value of a word in  $\text{Adh}(L)$ .

**Proposition 34.** *For all  $w \in \text{Adh}(L)$ ,*

$$\text{val}_S(w) = s_0 + \sum_{i=0}^{+\infty} \sum_{a < w[i]} r_{w[0, i-1]a}.$$

*Proof.* Let  $w \in \text{Adh}(L)$ . Using (6), we get that for all  $n \geq 1$ ,

$$\begin{aligned} \alpha_{w[0, n-1]} &= s_0 + \sum_{\substack{x < w[0, n-1] \\ x \in \text{Center}(L) \cap \Sigma^n}} r_x \\ &= s_0 + \sum_{i=0}^{n-1} \sum_{a < w[i]} \sum_{|y|=n-i-1} r_{w[0, i-1]ay} \\ &= s_0 + \sum_{i=0}^{n-1} \sum_{a < w[i]} r_{w[0, i-1]a}. \end{aligned}$$

Letting  $n$  tend to infinity in the latter equality, we get the expected result.  $\square$

The following proposition links together the framework of [LR02], where are mainly considered converging sequences of words, and the framework that has been developed in the present section to represent real numbers.

**Proposition 35.** *Let  $K$  be a language over a totally ordered alphabet  $(\Sigma, <)$  such that its prefix-closure  $\text{Pref}(K)$  satisfies Hypotheses (H1), (H2), and (H3), and let  $S = (\text{Pref}(K), \Sigma, <)$  be the abstract numeration system built on  $\text{Pref}(K)$ . If  $(w^{(n)})_{n \geq 0} \in K^{\mathbb{N}}$  is a sequence of words such that  $w^{(n)} \rightarrow w$ , then we have*

$$\lim_{n \rightarrow +\infty} \frac{\text{val}_S(w^{(n)})}{v_{q_0}(|w^{(n)}|)} = \alpha_w.$$

*Proof.* Let  $(w^{(n)})_{n \geq 0} \in K^{\mathbb{N}}$  be a sequence of words such that  $w^{(n)} \rightarrow w$ . Thanks to Lemma 5, this implies that  $\text{Pref}(w) \subseteq \text{Pref}(K)$ . For any  $\ell \geq 1$ , there exists  $N(\ell) \geq \ell$  such that for all  $n \geq N(\ell)$ ,  $w^{(n)}[0, \ell - 1] = w[0, \ell - 1]$ . Then in view of (7) and (8), for all  $\ell \geq 1$  and for all  $n \geq N(\ell)$ , we have

$$\left| \frac{\text{val}_S(w[0, |w^n| - 1])}{v_{q_0}(|w^n|)} - \frac{\text{val}_S(w^{(n)})}{v_{q_0}(|w^n|)} \right| \leq \frac{u_{q_0 \cdot w[0, \ell - 1]}(|w^n| - \ell)}{v_{q_0}(|w^n|)}.$$

Let  $\varepsilon > 0$ . By Hypothesis (H2), for all  $\ell \geq 1$ , there exists  $M(\ell) \geq \ell$  such that for all  $n \geq M(\ell)$ ,

$$\frac{u_{q_0 \cdot w[0, \ell - 1]}(|w^n| - \ell)}{v_{q_0}(|w^n|)} < r_{w[0, \ell - 1]} + \frac{\varepsilon}{2}.$$

By Hypothesis (H3), there exists  $k \in \mathbb{N}$  such that for all  $\ell \geq k$ ,  $r_{w[0, \ell - 1]} < \frac{\varepsilon}{2}$ . Then for all  $n \geq \max(N(k), M(k))$ , we have

$$\left| \frac{\text{val}_S(w[0, |w^n| - 1])}{v_{q_0}(|w^n|)} - \frac{\text{val}_S(w^{(n)})}{v_{q_0}(|w^n|)} \right| < \varepsilon.$$

□

To conclude this section, we recall some results from [BB97] interesting for our study.

**Proposition 36.** *If  $K$  is an infinite algebraic language over a totally ordered alphabet, then the minimal word of  $\text{Adh}(K)$  is ultimately periodic and can be effectively computed.*

**Definition 37.** Let  $K$  be a language over a totally ordered alphabet. The *minimal language* of  $K$ , denoted by  $\min(K)$  is the language of the smallest words of each length for the lexicographical ordering:

$$\min(K) = \{w \in K \mid \forall z \in K, |w| = |z| \Rightarrow w <_{\text{lex}} z\}.$$

**Proposition 38.** *If  $K$  is an infinite language such that  $K = \text{Center}(K)$ , then we have  $\min(K) = \text{Pref}(m_\varepsilon)$ .*

**Corollary 39.** *If  $K$  is an infinite algebraic language such that  $K = \text{Center}(K)$ , then  $\text{Pref}(m_\varepsilon)$  is a regular language.*

Of course, all these results can be adapted to the case of the maximal word of the adherence of a language.

Transposed to the context of this paper, these results can be related to syntactical properties of the endpoints of the intervals  $I_y$ , for  $y \in \text{Center}(L)$ .

**Corollary 40.** *Assume that the language  $L$  is algebraic. Then for all  $y \in \text{Center}(L)$ , the infinite words  $m_y$  and  $M_y$  are ultimately periodic.*

Notice that in general, there exist ultimately periodic representations that are not endpoints of any interval  $I_y$ , where  $y \in \text{Center}(L)$ . For instance, in the integer base 10 numeration system, we have that the representation of  $\frac{1}{3}$  is  $0.33333\ldots$  and  $\frac{1}{3}$  is not the endpoint of any interval of the form  $[\frac{k}{10^\ell}, \frac{k+1}{10^\ell}]$ , where  $\ell \geq 1$  and  $k \in [0, 10^\ell - 1]$ .

## 6. APPLICATIONS

In this section, we apply our techniques to three examples to represent real numbers in situations that were not settled in [LR02]. The first one shows how it can be easier to consider the prefix-closure of the language instead of the language itself.

**Example 41.** Consider again the language  $L = \{w \in \{a, b\}^* \mid |w|_a - |w|_b \leq 1\}$  of Example 2. This language is not prefix-closed. We have  $\text{Pref}(L) = \{a, b\}^*$ , which is of course a regular language. For the abstract numeration system  $S = (\text{Pref}(L), \{a, b\}, a < b)$ , the hypotheses (H1), (H2) and (H3) are trivially satisfied. More precisely, for all  $w \in \{a, b\}^*$ , we have  $r_w = 2^{-|w|-1}$ . Using the same notation as in Example 2, we have

$$\lim_{n \rightarrow +\infty} \frac{v_0(n-1)}{v_0(n)} = \frac{1}{2}.$$

Therefore, we represent the interval  $[\frac{1}{2}, 1]$ . For all  $\ell \geq 1$ ,  $\text{Center}(L) \cap \Sigma^\ell = \{a, b\}^\ell$  and the intervals corresponding to words of length  $\ell$  are exactly the intervals  $[\frac{k}{2^\ell}, \frac{k+1}{2^\ell}]$ , for any  $k \in [0, 2^\ell - 1]$ .

The second example illustrates the case of a non-regular language with a non-regular prefix-language.

**Example 42.** The *Dyck language* is the language

$$D := \{w \in \{a, b\}^* \mid |w|_a = |w|_b \text{ and } \forall u \in \text{Pref}(w), |u|_b \geq |u|_a\}$$

of the well-parenthesized words over two letters. Its (infinite) minimal automaton  $\mathcal{A}_D = (Q, q_0, \{a, b\}, \delta, \{q_0\})$  is represented in Figure 4. For each  $m \geq 0$ , define  $d_m = (a^m)^{-1}D = \{w \in \{a, b\}^* \mid a^m w \in D\}$  and  $d_{-1} = \emptyset$ , so that  $Q = \{d_m \mid m \geq 0\} \cup \{d_{-1}\}$ . Notice that in Figure 4, the states  $d_m$  are simply denoted by  $m$ .

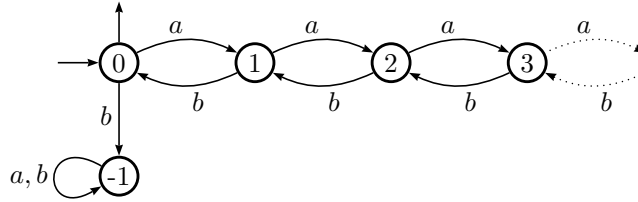


FIGURE 4. The minimal automaton of  $D$ .

It has been proved in [LG08] that for all  $m \geq 0$ ,

$$u_{d_m}(n) = \begin{cases} 0 & \text{if } n < m \text{ or } m \not\equiv n \pmod{2}, \\ \frac{m+1}{n+1} \binom{n+1}{\frac{n-m}{2}} & \text{if } n \geq m \text{ and } m \equiv n \pmod{2}. \end{cases}$$

By Stirling's formula, we get that for all  $m \geq 0$ ,

$$(9) \quad u_{d_{2m}}(2n) \sim \frac{2m+1}{\sqrt{\pi}} n^{-\frac{3}{2}} 4^n \quad (n \rightarrow +\infty),$$

$$(10) \quad u_{d_{2m+1}}(2n+1) \sim \frac{2(2m+2)}{\sqrt{\pi}} n^{-\frac{3}{2}} 4^n \quad (n \rightarrow +\infty).$$

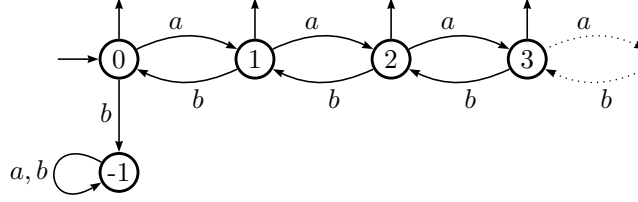
The Dyck language is not prefix-closed. Hence we consider the abstract numeration system  $S = (P, \{a, b\}, a < b)$  built on the language

$$P := \text{Pref}(D) = \{w \in \{a, b\}^* \mid \forall u \in \text{Pref}(w), |u|_b \geq |u|_a\}$$

of the prefixes of the Dyck words. The (infinite) minimal automaton of  $P$  is  $\mathcal{A}_P = (Q, q_0, \{a, b\}, \delta, F)$ . It is represented in Figure 5. Since the minimal automaton  $\mathcal{A}_P$  of  $P$  and the minimal automaton  $\mathcal{A}_D$  of  $D$  are nearly the same, we rename the states of  $\mathcal{A}_P$  by  $p_m := d_m$ . Hence the  $u_{d_m}$ 's denotes the complexity functions of  $\mathcal{A}_D$  and the  $u_{p_m}$ 's denotes the complexity functions of  $\mathcal{A}_P$ . By Proposition 13,  $\text{Adh}(P) = \text{Adh}(D)$  is uncountable and Hypothesis (H1) is satisfied.

Observe that for all  $m \geq 0$ ,

$$u_{p_m}(n) = \begin{cases} 2^n & \text{if } n \leq m, \\ 2u_{p_m}(n-1) - u_{d_m}(n-1) & \text{if } n > m. \end{cases}$$

FIGURE 5. The minimal automaton of  $\text{Pref}(D)$ .

Hence we get that for all  $m \geq 0$ ,

$$u_{p_m}(n) = \begin{cases} 2^n & \text{if } n \leq m, \\ 2^n - \sum_{i=m}^{n-1} u_{d_m}(i) 2^{n-i-1} & \text{if } n > m. \end{cases}$$

We have that for all  $m \geq 0$ ,

$$(11) \quad u_{p_m}(2n) \sim \frac{m+1}{\sqrt{\pi}} n^{-\frac{1}{2}} 4^n \quad (n \rightarrow +\infty),$$

$$(12) \quad u_{p_m}(2n+1) \sim v_{p_m}(2n) \sim \frac{2(m+1)}{\sqrt{\pi}} n^{-\frac{1}{2}} 4^n \quad (n \rightarrow +\infty),$$

$$(13) \quad v_{p_m}(2n+1) \sim \frac{4(m+1)}{\sqrt{\pi}} n^{-\frac{1}{2}} 4^n \quad (n \rightarrow +\infty).$$

We prove only (11) since the same techniques can be applied to obtain (12) and (13). Let us first show that for all  $m \geq 0$ , we have

$$(14) \quad \sum_{i=m}^{+\infty} u_{d_{2m}}(2i) 4^{-i} = 2 \quad \text{and} \quad \sum_{i=m}^{+\infty} u_{d_{2m+1}}(2i+1) 4^{-i} = 4.$$

We compute only the first sum, the second one can be treated in similar way. In view of (9) and [Bou07, Ch. V.4, Prop. 2], for all  $m \geq 0$ , we have

$$\sum_{i=n}^{+\infty} u_{d_{2m}}(2i) 4^{-i} \sim \frac{2m+1}{\sqrt{\pi}} \sum_{i=n}^{+\infty} i^{-\frac{3}{2}} \quad (n \rightarrow +\infty)$$

and the series

$$\sum_{i=m}^{+\infty} u_{d_{2m}}(2i) 4^{-i}$$

is convergent. Consequently, for all  $m \geq 0$ , the series

$$\sum_{i=m}^{+\infty} u_{d_{2m}}(2i) z^i$$

is uniformly convergent over  $\{z \in \mathbb{C} \mid |z| \leq \frac{1}{4}\}$  because for all  $q \geq p \geq m$ , we have

$$\sup_{|z| \leq \frac{1}{4}} \left| \sum_{i=p}^q u_{d_{2m}}(2i) z^i \right| \leq \sum_{i=p}^q u_{d_{2m}}(2i) 4^{-i}.$$

Then observe that for all  $m \geq 0$  and  $i \geq m$  such that  $i \equiv m \pmod{2}$ , we have

$$\begin{aligned} u_{d_m}(i) &= \text{Card}\{w^{(0)} b w^{(1)} b \dots b w^{(m)} \mid \forall j \in \llbracket 0, m \rrbracket, w^{(j)} \in D, \sum_{j=0}^m |w^{(j)}| = i - m\} \\ &= \sum_{\ell_0 + \dots + \ell_m = \frac{i-m}{2}} \left( \prod_{j=0}^m c_{\ell_j} \right) = \left[ z^{\frac{i-m}{2}} \right] \left( \sum_{n=0}^{+\infty} c_n z^n \right)^{m+1} \end{aligned}$$

where  $\mathcal{C}_n := u_{d_0}(2n) = \frac{1}{2n+1} \binom{2n+1}{n}$  is the  $n$ th Catalan number [GKP94] and  $[z^n]f$  is the coefficient of  $z^n$  in the power series  $f$ . It is well known that

$$\sum_{n=0}^{+\infty} \mathcal{C}_n z^n = \frac{1 - \sqrt{1-4z}}{2z}$$

for  $|z| < \frac{1}{4}$ . Hence we get that for all  $m \geq 0$ ,

$$\sum_{i=m}^{+\infty} u_{d_{2m}}(2i) z^i = z^m \left( \sum_{n=0}^{+\infty} \mathcal{C}_n z^n \right)^{2m+1} = \frac{(1 - \sqrt{1-4z})^{2m+1}}{2 \cdot 4^m z^{m+1}}$$

Therefore, we obtain the desired first sum of (14) by letting  $z$  tend to  $\frac{1}{4}$  in the corresponding formula. We now come back on (11). For all  $0 \leq m < n$ , we have

$$u_{p_{2m}}(2n) = 4^n - \frac{1}{2} \sum_{i=m}^{n-1} u_{d_{2m}}(2i) 4^{n-i} = \frac{1}{2} 4^n \sum_{i=n}^{+\infty} u_{d_{2m}}(2i) 4^{-i}$$

and

$$u_{p_{2m+1}}(2n) = 4^n - \frac{1}{4} \sum_{i=m}^{n-1} u_{d_{2m+1}}(2i+1) 4^{n-i} = \frac{1}{4} 4^n \sum_{i=n}^{+\infty} u_{d_{2m+1}}(2i+1) 4^{-i}.$$

Notice that  $\sum_{i=n}^{+\infty} i^{-\frac{3}{2}} \sim 2n^{-\frac{1}{2}}$ . Finally we obtain that for all  $m \geq 0$ ,

$$u_{p_{2m}}(2n) \sim \frac{2m+1}{\sqrt{\pi}} n^{-\frac{1}{2}} 4^n \quad \text{and} \quad u_{p_{2m+1}}(2n) \sim \frac{2m+2}{\sqrt{\pi}} n^{-\frac{1}{2}} 4^n,$$

proving (11).

Let us now verify that the language  $P$  satisfies our three hypotheses. From the previous reasoning, we get that for all  $m \geq 0$  and all  $\ell \geq 0$ ,

$$\lim_{n \rightarrow +\infty} \frac{u_{p_m}(n-\ell)}{v_{p_0}(n)} = (m+1) 2^{-\ell-1}.$$

For all  $w \in P$ ,  $r_w := (m_w + 1) 2^{-|w|-1}$  where  $m_w$  is defined by  $p_0 \cdot w = p_{m_w}$  and for all  $w \notin P$ ,  $r_w := 0$ . Hence Hypothesis (H2) is satisfied. Let now  $w \in \text{Adh}(D)$ . Observe that  $m_{w[0, \ell-1]} \leq \ell$  for all  $\ell \geq 1$ . Therefore, for all  $w \in \text{Adh}(D)$ , we have  $r_{w[0, \ell-1]} \leq (\ell+1) 2^{-\ell-1} \rightarrow 0$  as  $\ell \rightarrow \infty$  and Hypothesis (H3) is satisfied.

Since

$$\lim_{n \rightarrow +\infty} \frac{v_{p_0}(n-1)}{v_{p_0}(n)} = \frac{1}{2},$$

we represent the interval  $[\frac{1}{2}, 1]$ . We have  $\text{Center}(D) \cap \Sigma^\ell = P \cap \{a, b\}^\ell$ . Any word of  $P$  begins with  $a$ , so that  $I_a = [\frac{1}{2}, 1]$ . We have  $\text{Center}(D) \cap \Sigma^2 = \{aa, ab\}$  and  $I_a$  is partitioned into two subintervals:

$$I_{aa} = \left[ \frac{1}{2}, \frac{7}{8} \right] \quad \text{and} \quad I_{ab} = \left[ \frac{7}{8}, 1 \right]$$

Then  $\text{Center}(D) \cap \Sigma^3 = \{aaa, aab, aba\}$ . Thus  $I_{ab} = I_{aba}$  and  $I_{aa}$  is partitioned into two new subintervals

$$I_{aaa} = \left[ \frac{1}{2}, \frac{3}{4} \right], \quad I_{aab} = \left[ \frac{3}{4}, \frac{7}{8} \right], \quad I_{aba} = \left[ \frac{7}{8}, 1 \right].$$

Then  $\text{Center}(D) \cap \Sigma^4 = \{aaaa, aaab, aaba, aabb, abaa, abab\}$  and we get

$$\begin{aligned} I_{aaaa} &= \left[ \frac{1}{2}, \frac{21}{32} \right], \quad I_{aaab} = \left[ \frac{21}{32}, \frac{3}{4} \right], \quad I_{aaba} = \left[ \frac{3}{4}, \frac{27}{32} \right], \\ I_{aabb} &= \left[ \frac{27}{32}, \frac{7}{8} \right], \quad I_{abaa} = \left[ \frac{7}{8}, \frac{31}{32} \right], \quad I_{abab} = \left[ \frac{31}{32}, 1 \right]. \end{aligned}$$

As stated by Corollary 40, since the language  $D$  is algebraic, for all  $y \in \text{Center}(D)$ , the representations of the endpoints of the interval  $I_y$  are ultimately periodic. Let  $Q_x$  denotes the set of all the representations of  $x$ . We have  $Q_{\frac{1}{2}} = \{a^\omega\}$  and  $Q_1 = \{(ab)^\omega\}$ . Now let  $x \in (\frac{1}{2}, 1)$  be an endpoint of some interval, i.e.,  $x = \inf I_w = \sup I_z$  for some  $w, z \in \text{Center}(D) \cap \Sigma^\ell$  with  $\ell \geq 0$ . We have  $Q_x = \{\bar{w}(ab)^\omega, za^\omega\}$ , where  $\bar{w}$  is the smallest Dyck word having  $w$  as a prefix.

The third example illustrates the case of a generalized abstract numeration systems generating endpoints of the intervals  $I_y$  having no ultimately periodic  $S$ -representations. It also shows that our methods for representing reals generalize the ones involved to represent reals in the  $\frac{3}{2}$ -number system and by extension the rational base number systems as well.

**Example 43.** Consider the language  $L := L_{\frac{3}{2}}$  recognized by the deterministic automaton  $\mathcal{A} = (\mathbb{N} \cup \{-1\}, 0, \{0, 1, 2\}, \delta, \mathbb{N})$  where the transition function  $\delta$  is defined as follows:  $\delta(n, a) = \frac{1}{2}(3n + a)$  if  $n \in \mathbb{N}$  and  $a \in \{0, 1, 2\}$  are such that  $\frac{1}{2}(3n + a) \in \mathbb{N}$  and  $\delta(n, a) = -1$  otherwise. This language has been introduced and studied in [AFS08]. In particular, it has been shown that the automaton  $\mathcal{A}$  is the minimal automaton of  $L$ , that  $L$  is a non-algebraic prefix-closed language and that  $\text{Adh}(L)$  is uncountable. Moreover, no element of  $\text{Adh}(L)$  is ultimately periodic. The corresponding trim minimal automaton is depicted in Figure 6, where all states are final.

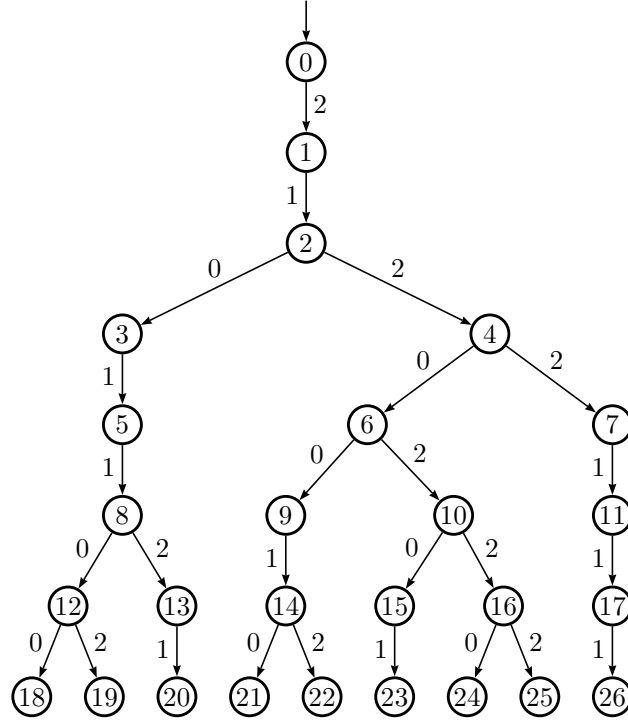


FIGURE 6. First levels of the trim minimal automaton of  $L_{\frac{3}{2}}$ .

Let  $(G_n)_{n \geq 0}$  be the sequence of integers defined by:

$$G_0 = 1 \quad \text{and} \quad \forall n \in \mathbb{N}, \quad G_{n+1} := \left\lceil \frac{3}{2} G_n \right\rceil.$$

From [AFS08], we find

$$u_0(0) = 1 \quad \text{and} \quad \forall n \in \mathbb{N}, \quad u_0(n+1) = G_{n+1} - G_n.$$



It has been shown in [AFS08] that for all  $n \geq 0$ ,  $G_n = \lfloor K \left(\frac{3}{2}\right)^n \rfloor$ , where  $K := K(3) = 1.6222705\dots$  is the constant discussed in [OW91, HH97, Ste03]. Consider now the abstract numeration system  $S = (L, \{0, 1, 2\}, 0 < 1 < 2)$  built on this language. From [AFS08], we know that for all  $w \in L$ ,

$$\text{val}_S(w) = \frac{1}{2} \sum_{i=0}^{|w|-1} w[i] \left(\frac{2}{3}\right)^{|w|-1-i}.$$

Consequently, for all  $w \in \text{Adh}(L)$ , we have

$$\text{val}_S(w) = \frac{1}{3K} \sum_{i=0}^{+\infty} w[i] \left(\frac{2}{3}\right)^i.$$

Now let us verify that  $L$  satisfies Hypothesis (H2) and (H3). Recall that, for all  $x \in L$ ,  $M_x$  (resp.  $m_x$ ) denotes the maximal (resp. minimal) word in  $\text{Adh}(L)$  for the lexicographic ordering having  $x$  as a prefix. We have that, for all  $x \in L$ ,

$$\begin{aligned} r_x = |I_x| &= \text{val}_S(M_x) - \text{val}_S(m_x) \\ &= \frac{1}{3K} \sum_{i=|x|}^{+\infty} (M_x[i] - m_x[i]) \left(\frac{2}{3}\right)^i \\ &= \frac{1}{3K} \left(\frac{2}{3}\right)^{|x|} \sum_{i=0}^{+\infty} (M_x[i+|x|] - m_x[i+|x|]) \left(\frac{2}{3}\right)^i \geq 0 \end{aligned}$$

and Hypothesis (H2) is satisfied. For all  $x \in L$ , since  $M_x[i] - m_x[i] \leq 2$  for all  $i \geq 0$ , we obtain from that

$$r_x \leq \frac{2}{K} \left(\frac{2}{3}\right)^{|x|} \rightarrow 0 \text{ as } |x| \rightarrow +\infty.$$

Therefore, if  $w \in \text{Adh}(L)$ , then  $\lim_{\ell \rightarrow +\infty} w[0, \ell - 1] = 0$  and Hypothesis (H3) is also satisfied.

#### OPEN PROBLEMS

- Find a necessary condition on any automaton recognizing a language  $L$  so that the corresponding  $\omega$ -language  $\text{Adh}(L)$  is uncountable.
- Let  $D_2$  be the Dyck language for two kinds of parentheses. It is well-known that for every algebraic language  $L$ , there exists a faithful sequential mapping  $f$  such that  $f(\text{Adh}(D_2)) = \text{Adh}(f(D_2)) = \text{Adh}(L)$ , see [BN80, Theorem 6] for details. Let  $S$  and  $T$  be abstract numeration systems built respectively on  $\text{Pref}(D_2)$  and  $\text{Pref}(L)$ . Give a mapping  $g$  such that the following diagram commutes.

$$\begin{array}{ccc} \text{Adh}(D_2) & \xrightarrow{f} & \text{Adh}(L) \\ \text{val}_S \downarrow & & \downarrow \text{val}_T \\ [s_0, 1] & \xrightarrow{g} & [t_0, 1] \end{array}$$

#### ACKNOWLEDGEMENTS

We thank Professor J.-P. Schneiders for fruitful discussions. This work was initiated during the post-doctoral stay of the second author at the University of Liege, thanks to a federal research allowance.

## REFERENCES

- [AFS08] S. Akiyama, C. Frougny, and J. Sakarovitch. Powers of Rationals Modulo 1 and Rational Base Number Systems. *Israel J. Math.*, 168:53–91, 2008.
- [BB97] J. Berstel and L. Boasson. The set of minimal words of a context-free language is context-free. *J. Comput. System Sci.*, 55(3):477–488, 1997.
- [BN80] L. Boasson and M. Nivat. Adherences of languages. *J. Comput. System Sci.*, 20(3):285–309, 1980.
- [Bou07] N. Bourbaki. *Fonctions d’une variable réelle*. Springer Berlin Heidelberg, 2007.
- [DT89] J.-M. Dumont and A. Thomas. Systèmes de numération et fonctions fractales relatifs aux substitutions. *Theoret. Comput. Sci.*, 65(2):153–169, 1989.
- [Eil74] S. Eilenberg. *Automata, Languages, and Machines*, volume A. Academic Press, New York, 1974. Pure and Applied Mathematics, Vol. 58.
- [GKP94] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete Mathematics*. Addison-Wesley Publishing Company, Reading, MA, second edition, 1994. A Foundation for Computer Science.
- [HH97] L. Halbeisen and N. Hungerbühler. The Josephus Problem. *J. Théor. Nombres Bordeaux*, 9(2):303–318, 1997.
- [LG08] M. Le Gonidec. On Complexity of Infinite Words Associated with Generalized Dyck Languages. *Theoret. Comput. Sci.*, 407(1-3):117–133, 2008.
- [Lot02] M. Lothaire. *Algebraic Combinatorics on Words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002.
- [LR01] P. B. A. Lecomte and M. Rigo. Numeration Systems on a Regular Language. *Theory Comput. Syst.*, 34(1):27–44, 2001.
- [LR02] P. B. A. Lecomte and M. Rigo. On the Representation of Real Numbers Using Regular Languages. *Theory Comput. Syst.*, 35(1):13–38, 2002.
- [Niv78] M. Nivat. Sur les ensembles de mots infinis engendrés par une grammaire algébrique. *RAIRO Inform. Théor.*, 12(3):259–278, v, 1978.
- [OW91] A. Odlyzko and H. Wilf. Functional Iteration and the Josephus Problem. *Glasgow Math. J.*, 33(2):235–240, 1991.
- [Sak03] J. Sakarovitch. *Éléments de théorie des automates*. Vuibert, Paris, 2003.
- [Ste03] R. Stephan. On a Sequence Related to the Josephus Problem. <http://arxiv.org/abs/math/0305348v1>, 2003.
- [SYZS92] A. Szilard, S. Yu, K. Zhang, and J. Shallit. Characterizing Regular Languages with Polynomial Densities. In *Mathematical Foundations of Computer Science 1992 (Prague, 1992)*, volume 629 of *Lecture Notes in Comput. Sci.*, pages 494–503. Springer, Berlin, 1992.

(E. Charlier, M. Rigo) INSTITUTE OF MATHEMATICS, UNIVERSITY OF LIÈGE, GRANDE TRAVERSE 12 (B 37), B-4000 LIÈGE, BELGIUM

*E-mail address:* {echarlier,M.Rigo}@ulg.ac.be

(M. Le Gonidec) LABORATOIRE IMATH, BAT U BUREAU 308 AVENUE DE L’UNIVERSIT - BP 20132 83957 LA GARDE CEDEX FRANCE TOLON, FRANCE

*E-mail address:* marion.le-gonidec@univ-tln.fr